

## О возможностях применения лингвостатистических подходов к анализу государственной научно-технологической политики

*В статье раскрыты методические возможности лингвостатистического подхода на базе закона Цифа для проведения анализа нормативно-правовых актов. Показано, что применение такого подхода к анализу законодательства позволяет оценить релевантность рангово-частотного распределения ключевых слов текста нормативно-правового акта целям этого документа (лингвистический анализ), а также оптимальность распределения слов в тексте закону Цифа, который определяет логико-семантическую структуру текста (статистический анализ). Тем самым подтверждена возможность применения этого метода в качестве аналитического инструмента науковедческого анализа государственной научно-технологической политики.*

**Ключевые слова:** государственная научно-технологическая политика, нормативно-правовой акт, законодательство, лингвостатистический подход, закон Цифа.

Государственная научно-технологическая политика осуществляется на основании действующей базы нормативно-правовых актов и при помощи управленческого исполнения соответствующих законодательно-правовых норм. Немаловажную роль в системе формирования и реализации государственной научно-технологической политики в Украине играет человеческий фактор [1], но качество профильных нормативно-правовых актов имеет не меньше значение.

Отсутствие науки в качестве реального приоритета развития современной украинской государственности не способствует разработке качественной правовой базы для проведения эффективной государственной политики. Действующие нормативно-правовые акты не обладают системностью, в первую очередь из-за отсутствия базового, кодификационного акта, задающего основные параметры правового регулирования в рамках самостоятельной законодательной отрасли [2].

Практическая актуальность проблем правового обеспечения законодательства в сфере научно-технологической де-

ятельности обуславливало применение преимущественно правоведческих методов к анализу нормативно-правовых актов, в том числе и учеными-науковедами. Правомерно, что законодательное обеспечение сферы научно-технологической деятельности стало актуальным направлением научных междисциплинарных исследований науковедения и правоведения и способствовало «становлению «правоведения науки» как одного из новых науковедческих направлений исследования науки, как особого социального института» [3, с. 39].

Актуальной проблемой науковедческого анализа законодательства становится применение методологических подходов, исследующих положения нормативно-правового акта не только по форме, но и по сути. А именно, методов, позволяющих исследовать соответствие нормативных положений актуальным направлениям научно-технологической деятельности и, в соответствии с проблемно-ориентированной оценкой научно-технологического потенциала, выявлять отображение в них (то есть в нормативных положени-

ях) механизмов государственного управления и регулирования в сфере науки и технологий.

Необходимой задачей также является разработка науковедческих подходов к анализу базы нормативно-правовых актов, одновременно как составной части государственной научно-технологической политики, так и самостоятельного направления научных исследований.

**Анализ последних исследований и публикаций.** Анализ государственной научно-технологической политики выступает предметом науковедческих исследований [4]. Основными методологическими подходами науковедческих исследований нормативно-правовых актов в области научно-технологической деятельности являются сравнительно-исторический анализ (источниковедение) [3], доктринальный анализ (соответствие содержания нормативного документа концептуальным, стратегическим доктринам развития науки) [5] и критический анализ содержания правовых норм [1; 6]. Перечисленные научные подходы в основном основаны на методе сравнительно-правового анализа законодательства [2; 7; 8], который определяет соответствие нормативно-правовых актов предмету правового регулирования, а не его содержательности («букву закона, а не его дух» [9, с. 144]). Для исследований текстов нормативно-правовых актов (содержательности) в сфере научно-технологической деятельности применяются методы логико-семантического и историко-правового анализа [10], а также контент-анализа [11].

При помощи логико-семантического метода исследуют формализацию законодательного языка и соответствие слов их юридическому контексту [12]. Метод контент-анализа позволяет определить соответствие положений нормативного текста смысловому содержанию направлений государственной политики в определенной сфере деятельности [13].

Методологической основой контент-анализа при исследовании текстов является лингвостатистический подход, использующий закономерности

рангово-частотного распределения слов (закон Ципфа). Лингвостатистические исследования текстов с использованием программных (математических) и лексико-семантических способов проводятся в рамках отдельного научного направления – лингвостатистики [14]. Лингвостатистический подход для анализа текстов используется в интернет-пространстве (веб-технологии) и в языковедении. Лингвостатистические веб-технологии, или метод SEO оптимизации (*Search Engine Optimization*), используются как в качестве алгоритмов поисковых систем (*Google, Yandex, Rambler*), так и для рационализации текстов, составляющих контентное содержание сайтов [15]. В языковедении статистические закономерности рангово-частотного распределения слов в тексте используются для анализа особенностей авторского текста литературных произведений, составления частотных словарей и тезауруса, литературных машинных переводов [16–18].

Малоизученным остается применение лингвостатистического подхода для анализа текста нормативно-правовых актов, которое представляется перспективным с точки зрения возможностей применения веб-технологий.

**Целью** работы является обоснование применения лингвостатистического подхода как науковедческого метода исследования государственной научно-технологической политики.

Научно-технологическая политика является предметом науковедческих исследований, на что указывал еще Геннадий Михайлович Добров в своем обосновании научного направления «науковедение» «как комплексного исследования и теоретического обобщения опыта функционирования социальных систем в науке с целью обоснования научно-технической политики» [19, с. 26]. Основы государственной научно-технологической политики определяются законодательством большинства государств через систему нормативно-правовых актов, имплементация которых

предполагает реализацию механизмов государственной политики в сфере научно-технологической деятельности.

**Законодательством** называется совокупность всех нормативно-правовых актов, которые действуют в государстве. В зависимости от государственного статуса субъекта государственного управления это понятие, в соответствии с толкованием Конституционного Суда Украины, употребляется в трех значениях [20]:

1) **узком**, как система законов, которая формируется Верховной Радой Украины как единым органом законодательной власти. В эту систему также включены международные соглашения, которые ратифицированы Верховной Радой Украины;

2) **широком**, как система законов и других нормативно-правовых актов, которые принимаются высшими органами власти – Верховной Радой Украины, Президентом Украины и Кабинетом Министров Украины (постановления, указы, распоряжения);

3) **всеобщем**, как система законов и других нормативно-правовых актов, которые принимаются не только высшими органами власти, но и центральными и местными органами исполнительной власти и самоуправления.

Содержание законодательства (структура нормативно-правовых актов, системность и качество правовых норм) влияет не только на эффективность государственной научно-технологической политики, но и на состояние и тенденции развития науки и технологий в стране. Объектом науковедческих исследований является не только законодательство в сфере научно-технологической деятельности (профильное), но и общее законодательство. Это диктуется внутренней сущностью науки – междисциплинарной природой научных исследований и мультифункциональной областью применения научных результатов. Влияние состояния не только профильного, но и общего законодательства на сферу научно-технологической деятельности подтверждается тем, что:

- изменение профильных нормативно-правовых актов приводит к изменению базовых условий осуществления научно-технологической деятельности, что приводит к кардинальным изменениям структуры сферы науки и технологий;

- изменение общего законодательства (налогового, кредитно-финансового законодательства, законов о государственном бюджете и др.) приводит к кардинальным и локальным изменениям структуры сферы науки и технологий;

- снижение государственной приоритетности научно-технологических факторов в общем законодательстве (например, в законодательстве в сфере национальной безопасности) приводит к деградации ряда высокотехнологических отраслей промышленности – машиностроения, ракетно-космической отрасли, оборонно-промышленного комплекса.

При науковедческом анализе государственной научно-технологической политики законодательство одновременно выступает как причиной, так и следствием осуществления государственного управления наукой. Отсутствие в науковедении собственной методологии исследований нормативно-правовых актов компенсировалось применением юридических, философских, политических и социологических методов анализа, которые используются для решения отдельных научных задач, связанных с механизмами формирования и реализации государственной научно-технологической политики. Применение методов исследования из других научных дисциплин является частью методологического богатства науковедения, поскольку «на основании этого науковедение вырабатывает свои собственные науковедческие факты, методы и концепции» [19, с. 26].

Лингвостатистический подход на базе закона Ципфа основан на количественной оценке информации, что позволяет использовать для аналитического исследования текста программные средства, в том числе современные веб-технологии. В ходе многочисленных исследований было доказано, что закону

Ципфа хорошо подчиняются количественные характеристики литературных произведений определенного жанра (частота употребления ключевых слов и выражений) и плохо – произвольные лексические отрывки, не обладающие сюжетной однородностью [21; 22].

Основные критерии целесообразности применения лингвостатистического подхода к анализу текстов, а именно, **целостность, завершенность и однородность стиля**, вполне согласуются с правилами

нормотворческой деятельности [23]. Это означает, что текст любого нормативно-правового акта может быть исследован методиками, основанными на использовании статистических закономерностей Ципфа.

Закон Ципфа отображает закономерность «ранг-частота» распределения словоформ (грамматических форм слов) в любом естественном тексте. Аналитически эта закономерность описывается в виде закона Ципфа–Мандельброта [24]:

$$f(r) = \frac{C}{(B+r)^\gamma}$$

где  $f(r)$  – частота употребления одной словоформы в тексте нормативно-правового акта;

$r$  – ранг словоформы в списке ранжирования частот по убыванию;

$C$  – постоянная величина, которая определяется через соотношение с количеством слов в тексте ( $N$ ) как

$$\frac{C}{N} = const$$

Величина этой константы различается для разных групп языков: для английского и французского языков она равна 0,1; для славянских языков – 0,06–0,08.

$\gamma$  – показатель степени рангово-частотного распределения словоформ, который варьируется в зависимости от стиля языка текста (для нормативно-правовых актов – официально-деловой стиль) и по своему значению близок к единице.  $B$  – малая постоянная величина, отличающаяся от нуля в области высокочастотных слов (словоформ).

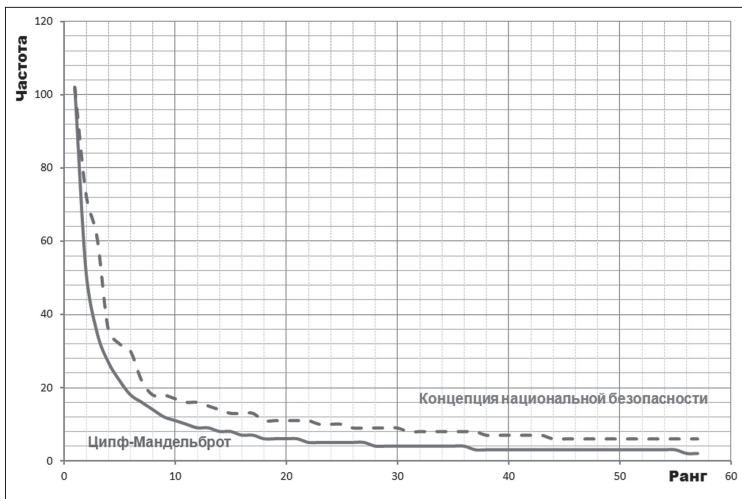
Показатели  $\gamma$  и  $B$  были введены Мандельбротом для улучшения соответствия эмпирических (реальных) и теоретических данных рангово-частотного распределения слов в тексте. При  $B=0$  аналитическая зависимость Ципфа–Мандельброта представляется как модифицированный закон Ципфа.

Графически зависимость распределения частоты словоформы от ее ранга отображается кривой, близкой к гиперболе (рис.).

Верхняя кривая на графике отображает реальное распределение частоты употреблений словоформ в тексте, нижняя – представляет ее теоретическую взаимозависимость согласно закону Ципфа. Не прибегая сейчас к формальным (математическим) методам оценки соответствия рангово-частотной зависимости словоформ в тексте Концепции национальной безопасности Украины теоретическому распределению Ципфа, на основании графического представления на рис. мы уже можем подтвердить возможность использования лингвостатистического подхода к анализу текста этого нормативно-правового акта.

Лингвостатистический подход позволяет выделить семантическое ядро текста, которое составляют ключевые слова и словосочетания. Относительно нормативно-правовых актов содержание такого семантического ядра представляет собой сущность государственной политики [13].

Ключевые слова, которые определяются при лингвостатистическом анализе



**Рис. Рангово-частотное распределение первых 58 словоформ в тексте Концепции национальной безопасности Украины [25]**

текста, — это приведенные к каноническому (словарному) виду словоформы, которые совместно с другими ключевыми словосочетаниями представляют содержание текста. Ключевое словосочетание — это смысловое и грамматическое словосочетание, состоящее из одного или нескольких ключевых слов. Как правило, словосочетания состоят из имен существительных и подчиненных им прилагательных и/или имен существительных.

Алгоритм выделения ключевых слов и словосочетаний нормативно-правового акта состоит в следующем.

1. Из текста нормативно-правового акта исключаются стоп-слова (т. н. «шумовые» слова): предлоги, суффиксы, причастия, междометия, цифры, частицы и другие неинформативные термины.

2. К оставшимся словам текста применяется **лемматизация** — процедура приведения словоформ к словарному виду.

3. В результате проведения семантического анализа составляется частотный словарь текста, в котором лемматизированные словоформы располагаются в порядке (по рангам) убывания частоты употребления [26].

4. Осуществляется рангово-частотный анализ слов из полученного частотного словаря, в результате которого все

слова (лемматизированные словоформы) сводятся в две группы — высокочастотных и низкочастотных.

5. Осуществляется выбор ключевых слов и словосочетаний из высокочастотной группы слов, соответствующих тематике (направленности) нормативно-правового акта. Определяется удельный вес ключевых слов относительно объема текста. Как правило, ключевые слова и ключевые словосочетания присутствуют в названии нормативно-правового акта, а также составляют семантическое ядро преамбулы закона (в случае наличия преамбулы). Анализ низкочастотных слов представляет собой отдельный предмет лингвостатистических исследований.

Процедура выделения ключевых слов и словосочетаний может осуществляться тремя способами: вручную, при помощи специализированных статистических программ (например, пакета статистических программ STATISTICA) или веб-технологий. Наиболее простым способом является проведение лингвостатистической обработки текста при помощи веб-технологий, однако их использование для анализа текстов нормативно-правовых актов имеет свои преимущества и недостатки. К преимуществам



ществам (достоинствам) веб-технологий относятся:

– существенная экономия времени на предварительную обработку текста на русском и английском языках;

– лингвостатистическая обработка текстов объемом до 15000 слов / 100000 знаков (примерно 30 страниц формата А4);

– автоматическое выстраивание рангово-частотного распределения слов;

– выбор бесплатных программных средств из сети Интернет.

Недостатки применения веб-технологий для анализа текста нормативно-правового акта являются продолжением их достоинств:

– отсутствие доступных сетевых программ для лингвостатистического анализа текстов на украинском языке (на момент написания статьи только сайт [<http://advego.ru/text/seo/>] предлагал такую возможность, однако из-за нечеткости сопоставления словоформ с их словарным видом требовал ручного пересчета отдельных значений рангово-частотного распределения слов в тексте);

– методическая нечеткость критериев выделения зоны ключевых слов и словосочетаний (корпоративная закрытость таких алгоритмов на поисковых системах в связи с их коммерческой ценностью);

– применение разных сетевых программ дает разные результаты рангово-частотного распределения ключевых слов в одном и том же тексте.

С учетом перечисленных недостатков применения веб-технологий целесообразно проведение лингвостатистического анализа текста нормативно-правового акта с использованием комбинирования способов.

Применительно к государственной научно-технологической политике анализ профильного законодательства может рассматриваться как зависимость целей научно-технологической политики от содержания нормативно-право-

вых актов. Цели политики в этом случае выступают ранговым показателем (уровнем приоритетности), а законодательство – частотой упоминания о целях (уровнем обеспечения). Результаты лингвостатистического анализа общего законодательства могут рассматриваться как распределение усилий государства в разных сферах жизнедеятельности общества для достижения целей государственной научно-технологической политики. Для адекватного определения ключевых слов, отображающих цели научно-технологической политики в нормативно-правовых актах общего законодательства, целесообразно разработать специальный тезаурус, содержащий науковедческие понятия.

**Выводы.** Таким образом, применение лингвостатистического подхода к анализу законодательства позволяет оценить: релевантность рангово-частотного распределения ключевых слов текста нормативно-правового акта его целям (лингвистический анализ), а также оптимальность распределения слов в тексте закону Ципфа, который определяет логико-семантическую структуру текста (статистический анализ). Тем самым подтверждается возможность применения этого метода в качестве аналитического инструмента науковедческого анализа государственной научно-технологической политики.

Практическая ценность лингвостатистического подхода на базе закона Ципфа для актуальных задач науковедения будет изучена по результатам дальнейших исследований нормативно-правовых актов в сфере научно-технологической деятельности. Результаты проведенных исследований лягут в основу предложений относительно законодательных изменений основного закона о научной и научно-технической деятельности в Украине, в том числе на предмет его кодификации.

1. *Попович О. С.* Науково-технологічна та інноваційна політика: основні механізми формування та реалізації : моногр. / О. С. Попович ; під ред. д-ра екон. наук,

проф. Б. А. Маліцького. – К. : Фенікс, 2005. – 226 с.

2. *Правовая* политика и пути совершенствования правотворческой деятельнос-

- ти в Российской Федерации : монография / Варламова Н. В., Лапаева В. В., Лукашева Е. А., Малахов В. П. и др. ; Под ред. Соколова Н. С. – М. : Изд-во РУДН, 2006. – 542 с.
3. *Аллахвердян А. Г.* Динамика научных кадров в советской и российской науке: сравнительно-историческое исследование / А. Г. Аллахвердян. – М. : Изд-во «Когито-Центр», 2014. – 263 с.
  4. *Актуальні питання методології та практики науково-технологічної політики / Б. А. Малицький, І. О. Булкін, І. Ю. Єгоров, Л. П. Кавуненко, Г. І. Калитич, всього 8 осіб / Під ред. Б. А. Малицького.* – К. : УкрІНТЕІ. – 2001. – 204 с.
  5. *Малицький Б. А.* Прикладне наукознавство / Б. А. Малицький. – К. : Фенікс, 2007. – 464 с.
  6. *Дежина И. Г.* Государственное регулирование науки в России / И. Г. Дежина / Под. ред. Н. И. Ивановой. – М. : Магистр, 2008. – 430 с.
  7. *Пахомова А. О.* Загальна характеристика законодавства у сфері проведення науково-дослідних робіт / А. О. Пахомова // Вісник Академії адвокатури України. – 2011. – Число 2 (21). – С. 52–58.
  8. *Селіванов А. О.* Наука і закон. Перший досвід системного аналізу законодавства у сфері науки і науково-технічної діяльності / А. О. Селіванов. – К. : ЛОГОС, 2002. – 263 с.
  9. *Ткачук А.* Законодавча техніка : навч. посібник / Анатолій Ткачук. – К. : ІКЦ «Легальний статус», 2011. – 268 с.
  10. *Мосьондз С. О.* Адміністративно-правові засади державної політики у сфері науки в Україні : автореф. дис. на здобуття наук. ступеня д-ра юрид. наук: спец. 12.00.07 «Адміністр. право і процес; фінансове право; інформаційне право» / Мосьондз Сергій Олександрович ; Відкр. міжнар. ун-т розв. людини «Україна». – К., 2012. – 37 с.
  11. *Теоретичні та методологічні проблеми розробки і реалізації управлінських стратегій : монографія / За заг. ред. В. М. Князева.* – К. : НАДХ, 2008. – 240 с.
  12. *Юркевич О. М.* Нульові поняття у чинному законодавстві України: логіко-семантичний аналіз / О. М. Юркевич, Н. Ю. Філатова // Вісник Харківського Національного університету внутрішніх справ. – 2009. – Вип. 45. – С. 14–22.
  13. *Лузан В. С.* Контент-анализ основных нормативных правовых документов, обеспечивающих реализацию государственной культурной политики (федеральный и региональный аспекты) / В. С. Лузан // Журнал Сибирского федерального университета. Сер. 4. Гуманитарные науки. – 2011. – № 3. – С. 342–362.
  14. *Лінгвостатистика (курс лекцій) / Укл.: П. Д. Варбанець, Н. А. Якімова.* – Одеса, Одеський Нац. ун-т ім. І. І. Мечникова. – 2013. – 49 с.
  15. *Энциклопедия SEO: Поисковая оптимизация от «А до Я» [Интернет-ресурс].* Режим доступа: <http://www.seobuilding.ru>.
  16. *Бук С.* Слов'янський досвід укладання частотних словників мови письменника / С. Бук // Проблеми слов'янознавства. – 2011. – Вип. 60. – С. 217–224.
  17. *Волошиновська І. А.* Сильова, тематична й авторська атрибуція наукових і художніх текстів (на матеріалі англійської, німецької та української мов) : автореф. дис. на здобуття наук. ступеня канд. філолог. наук : спец. 10.02.15 «Загальне мовознавство» / Волошиновська Ірина Анатоліївна; Нац. педагог. ун-тет ім. М. П. Драгоманова. – К., 2013. – 21 с.
  18. *Ягунова Е. В.* Эксперимент и вычисления в анализе ключевых слов художественного текста / Е. В. Ягунова // Сборник научных трудов кафедры иностранных языков и философии ПНЦ УрО РАН. Философия языка. Лингвистика. Лингводидактика. – Пермь, 2010. – Вып. 1. – С. 85–91.
  19. *Добров Г. М.* Наука о науке: начала науковедения / Г. М. Добров. – [3-е изд. доп. и перераб.] – К. : Наук. думка, 1989. – 304 с.
  20. *Сущенко В.* Тлумачення терміна «законодавство» Конституційним Судом України / Володимир Сущенко, Олена Царенко // Юридичний журнал. – 2003. – № 5. – С. 128–130.
  21. *Ягунова Е. В.* Динамические частотные характеристики как основа для структурного описания разнородных лингвистических объектов / Е. В. Ягунова,

- Д. В. Ландэ // Труды 14-й Всероссийской научной конференции. Электронные библиотеки: перспективные методы и технологии, электронные коллекции. – RCDL-2012, Переславль-Залесский, Россия, 15–18 октября 2012 г. – С. 196–205.
22. *Фирстов В. Е.* Из истории стенографии и интерпретации закона Ципфа-Мандельброта / В. Е. Фирстов, Р. А. Иванов // Ярославский педагогический вестник. – 2013. – № 4, Том III (Естественные науки). – С. 7–13.
23. *Постанова* Кабінету Міністрів України від 06.09.2005 року № 870 «Про затвердження Правил підготовки проєктів актів Кабінету Міністрів України» (Зі змінами, внесеними згідно з Постановою КМ № 1212 (1212-2009-п ) від 18.11.2009) // Офіційний вісник України від 23.09.2005. – 2005 р., № 36, стор. 45, ст. 2200.
24. *Арапов М. В.* О смысле ранговых распределений / М. В. Арапов, Е. Н. Ефимова, Ю. А. Шрейдер // Научно-техническая информация. – Сер. 2. – 1975. – № 1. – С. 9–20; Дунаев В. О ранговых распределениях в классификации / Вадим Дунаев // Научно-техническая информация. Сер. 2. – 1984 [Электронный ресурс]. – Режим доступа: <http://dunaevv1.narod.ru/other/classrunk.htm>; Орлов Ю. К. Динамика ранговых распределений и проблемы статистики редких событий / Ю. К. Орлов ; в кн. : Математическое описание ценозов и закономерности технетики. Вып. 1 и вып. 2. «Ценологические исследования» ; Ред. и сост. Б. И. Кудрин. – Абакан : Центр системных исследований, 1996. – С. 79–93.
25. *Про Концепцію* (основи державної політики) національної безпеки України // Відомості Верховної Ради України. – 1997. – № 10. – Ст. 85.
26. *Бук С.* Частотний словник офіційно-ділового стилю: принципи укладання та статистичні характеристики лексики / Соломія Бук // Лінгвістичні студії. – Вип. 14. – 2006. – С. 184–188.

Получено 16.05.2015

С. Г. Бублик

### Про можливості застосування лінгвостатистичних підходів до аналізу державної науково-технологічної політики

*У статті розкрито методичні можливості лінгвостатистичного підходу на базі закону Ципфа для проведення аналізу нормативно-правових актів. Показано, що застосування такого підходу до аналізу законодавства дозволяє оцінити релевантність рангово-частотного розподілу ключових слів тексту нормативно-правового акту цілям цього документа (лінгвістичний аналіз), а також оптимальність розподілу слів у тексті закону Ципфа, який визначає логіко-семантичну структуру тексту (статистичний аналіз). Тим самим підтверджено можливість застосування цього методу як аналітичного інструменту наукознавчого аналізу державної науково-технологічної політики.*

**Ключові слова:** державна науково-технологічна політика, нормативно-правовий акт, законодавство, лінгвостатистичний підхід, закон Ципфа.